

- KLEVYTSKA, A. M., L. B. PRICE, J. M. SCHUPP, P. L. WORSHAM, J. WONG and P. KEIM (2001): Identification and characterization of variable-number tandem repeats in the *Yersinia pestis* genome. *J. Clin. Microbiol.* **39**: 3179–3185.
- LINCOLN, S., M. DALY and E. LANDER (1992): Constructing genetic maps with MAPMAKER/EXP 3.0. Whitehead Institute technical report, 3<sup>rd</sup> edition. Whitehead Institute, Cambridge, MA.
- MACAVOY, E. S., L. M. MCGIBBON, J. P. SAINSBURY, H. LAWRENCE, C. A. WILSON, C. H. DAUGHERTY and G. K. CHAMBERS (2007): Genetic variation in island populations of tuatara (*Sphenodon* spp) inferred from microsatellite markers. *Conserv. Genet.* **8**: 305–318.
- MAGNUSSON, P. K. E., E. WILANDER and U. GYLLENSTEN (1996): Analysis of loss of heterozygosity in microdissected tumor cells from cervical carcinoma using fluorescent dUTP labeling of PCR products. *BioTechniques* **21**: 844–847.
- NAGARAJU, J., M. KATHIRVEL, R. R. KUMAR, E. A. SIDDIQ and S. E. HASNAIN (2002): Genetic analysis of traditional and evolved Basmati and non-Basmati rice varieties by using fluorescence-based ISSR-PCR and SSR markers. *Proc. Natl. Acad. Sci. USA* **99**: 5836–5841.
- NGUYEN, H. T. and X. WU (2005): Molecular marker systems for genetic mapping. In: *The handbook of plant genome mapping: genetic and physical mapping*, eds. K. MEKSEM and G. KAHL. WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany. 23–52.
- OTTEWELL, K. M., S. C. DONNELLAN, G. F. MORAN and D. C. PATON (2005): Multiplexed microsatellite markers for the genetic analysis of *Eucalyptus leucoxylon* (Myrtaceae) and their utility for ecological and breeding studies in other *Eucalyptus* species. *J. Hered.* **96**: 445–451.
- PASHLEY, C. H., J. R. ELLIS, D. E. MCCAULEY and J. M. BURKE (2006): EST databases as a source for molecular markers: lessons from *Helianthus*. *J. Hered.* **97**: 381–388.
- POWELL, W., G. C. MACHRAY and J. PROVAN (1996): Polymorphisms revealed by simple sequence repeats. *Trend. Plant Sci.* **1**: 215–222.
- SCHLÖTTERER, C. (2004): The evolution of molecular markers—just a matter of fashion? *Nat. Rev. Genet.* **5**: 63–69.
- SCHUELKE, M. (2000): An economic method for the fluorescent labeling of PCR fragments. *Nat. Biotech.* **18**: 233–234.
- TAUZ, D. (1989): Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Res.* **16**: 6463–6471.
- THAMARUS, K., K. GROOM, J. MURRELL, M. BYRNE and G. F. MORAN (2002): A genetic linkage map for *Eucalyptus globulus* with candidate loci for wood, fibre and floral traits. *Theor. Appl. Genet.* **104**: 379–387.
- WILLIAMS, R. D. and J. A. DEWOODY (2003): Fluorescent dUTP helps characterize 10 novel tetranucleotide microsatellites from an enriched salamander (*Ambystoma texanum*) genomic library. *Mol. Ecol. Notes* **4**: 17–19.
- WOOLBRIGHT, S. A., S. P. DIFAZIO, T. YIN, G. D. MARTINSEN, X. ZHANG, G. J. ALLAN, T. G. WHITHAM and P. KEIM (2008): A dense linkage map of hybrid cottonwood (*Populus fremontii* × *P. angustifolia*) contributes to long-term ecological research and comparison mapping in a model forest tree. *Heredity* **100**: 59–70.
- YASODHA, R., R. SUMATHI, P. CHEZHIAN, S. KAVITHA and M. GHOSH (2008): *Eucalyptus* microsatellites mined *in silico*: survey and evaluation. *J. Genet.* **87**: 21–25.

## Two-dimensional penalized splines via Gibbs sampling to account for spatial variability in forest genetic trials with small amount of information available

By E. P. CAPP<sup>1,\*</sup>, M. LSTIBUREK<sup>2</sup>, A. D. YANCHUK<sup>3,4</sup> and Y. A. EL-KASSABY<sup>4</sup>

(Received 14<sup>th</sup> January 2010)

<sup>1</sup>) Instituto Nacional de Tecnología Agropecuaria (INTA), Instituto de Recursos Biológicos, De Los Reseros y Dr. Nicolás Repetto s/n, 1686, Hurlingham, Buenos Aires, Argentina.

<sup>2</sup>) Department of Dendrology and Forest Tree Breeding, Faculty of Forestry and Wood Sciences, Czech University of Life Sciences Prague, Kamycka 129, 165 21 Praha 6, Czech Republic.

<sup>3</sup>) British Columbia Forest Service, Tree Improvement Branch, Victoria, British Columbia, V8W 9C2 Canada.

<sup>4</sup>) Department of Forest Sciences, Faculty of Forestry, University of British Columbia, 2424 Main Mall, Vancouver, British Columbia, V6T 1Z4 Canada.

\*) Corresponding author: EDUARDO PABLO CAPP. Bosques Cultivados Instituto de Recursos Biológicos, Centro de Investigación en Recursos Naturales, Instituto Nacional de Tecnología Agropecuaria, De Los Reseros y Dr. Nicolás Repetto s/n, 1686 Hurlingham, Buenos Aires, Argentina. Phone: +54 11 4621 0433, Fax: +54 11 4621 0433. E-mail: [ecappa@cni.inta.gov.ar](mailto:ecappa@cni.inta.gov.ar)

### Abstract

Spatial environmental heterogeneity are well known characteristics of field forest genetic trials, even in small experiments (<1 ha) established under seemingly uniform conditions and intensive site management. In such trials, it is commonly assumed that any simple type of experimental field design based on randomization theory, as a completely randomized design (CRD), should account for any of the minor site variability. However, most published results indicate that in these types of trials harbor a large component of the spatial variation which commonly resides in the error term. Here we applied a two-dimensional smoothed surface in an individual-tree mixed model, using tensor product of linear, quadratic and cubic B-spline bases with different and equal number of knots for rows and columns, to account

for the environmental spatial variability in two relatively small (i.e., 576 m<sup>2</sup> and 5,705 m<sup>2</sup>) forest genetic trials, with large multiple-tree contiguous plot configurations. In general, models accounting for site variability with a two-dimensional surface displayed a lower value of the deviance information criterion than the classical RCD. Linear B-spline bases may yield a reasonable description of the environmental variability, when a relatively small amount of information available. The mixed models fitting a smoothed surface resulted in a reduction in the posterior means of the error variance ( $\sigma_e^2$ ), an increase in the posterior means of the additive genetic variance ( $\sigma_a^2$ ) and heritability ( $h_{HT}^2$ ), and an increase of 16.05 % and 46.03 % (for parents) or 11.86 % and 44.68 % (for offspring) in the accuracy of breeding values, respectively in the two experiments.

**Key words:** genetics trials, spatial variability, tensor product of B-spline, error variance, genetic parameters, accuracy.

## Introduction

Spatial environmental heterogeneity are well known characteristics of field forest genetic trials (e.g., DUTKOWSKI et al., 2006; ZAS, 2006; CAPPA and CANTET, 2007; YE and JAYAWICKRAMA, 2008; FINLEY et al., 2009). This is true even in small experiments (<1 ha) established under seemingly uniform conditions and intensive site management (WOODS et al., 1995; SAENZ-ROMERO et al., 2001; JOYCE et al., 2002). In such trials, it is commonly assumed that the amount (magnitude and direction) of site variability is minimal, with only small random micro-site variations being present among experimental units (i.e., plots, trees or plants). Although the completely randomized design (CRD) is generally considered adequate for revealing important differences among the studied sources of variation (LOO-DINKINS 1992), in most cases the amount of variation residing in the “error term” could be rather high (e.g., EL-KASSABY and PARK, 1993; REHFELDT, 1995; KRAKOWSKI et al., 2005; ST. CLAIR, 2006). While the major advantage of the CRD, aside from being the easier to establish, is the simplicity of its analysis, such simple experimental design is unlikely to account for the majority of environmental variation. To reduce the impact of environmental variability, statisticians, crop and tree breeders have designed or adopted more efficient experimental layouts. Randomized complete block (RCB) or incomplete blocks designs, attempt an *a priori* to separate the site’s heterogeneity into homogeneous blocks. However, setting such an assumption is often unrealistic or weak as well, as two most-distant measurements taken within the same block should in theory share the same variance, while two close measurements of the neighboring trees on the border of two blocks are assumed to vary by a different magnitude. The magnitude of this discrepancy increases as the size of experiment increases. One strategy to resolve this issue is to minimize the size of the experiment, by considering the desired power of the experiment, and limiting the size to a minimum. Unfortunately, in most cases, even under the most efficient experimental layout, the spatial heterogeneity is unknown during the establishment phase and is only revealed at the evaluation stage. Thus, it is necessary to

model such variability *a posteriori* within the model of evaluation.

Spatial models permit modeling site heterogeneity by including two main components; namely, the “local trend” or small-scale and the “global trend” or large-scale variations (GRONDONA et al., 1996). Several *a posteriori* approaches have been developed and applied to forest genetic trials to more accurately account for site heterogeneity. The impact of small-scale spatial heterogeneity is classically accounted for through the inclusion of a random spatially correlated structure into the model. Such a residual matrix is expressed as a Kronecker product of the first order autoregressive residuals for rows and columns (GILMOUR et al., 1997). Additionally, in forest genetic and other trials, the small-scale spatial variability has been modeled with either the nearest neighbor techniques (MAGNUSSEN, 1990; ANEKONDA and LIBBY, 1996; JOYCE et al., 2002; KROON et al., 2008), or kriging (HAMANN et al., 2002; ZAS, 2006). Some approaches that account for large-scale continuous spatial variation have been modeled through post-blocking (ERICSSON, 1997; LOPEZ et al., 2002; GEZAN et al., 2006; KROON et al., 2008), or the inclusion of spatial coordinates expressed as either classification variables such as polynomials (THOMSON and EL-KASSABY, 1988; FEDERER, 1998; SAENZ-ROMERO et al., 2001) or covariables or smoothing splines (GILMOUR et al., 1997; VERBYLA et al., 1999). GILMOUR et al. (1997), in agricultural trials, and COSTA e SILVA et al. (2001) and DUTKOWSKI et al. (2002) in forest tree field experiments, recommended modeling the small-scale variation by fitting separable two-dimensional autoregressive residuals and the large-scale variation (global) in one dimension through fixed or random classification variables (COSTA e SILVA et al., 2001), or included fixed effects of spatial coordinates as either polynomials or cubic smoothing splines (DUTKOWSKI et al., 2002). However, the fit of a global trend plus autoregressive residuals were not successful or yielded little or no improvement, and thus DUTKOWSKI et al. (2006) recommended retaining the design terms in the spatial model. Additionally, a large portion of the global trend is usually present in two dimensions and non-stochastic functions as such as polynomials (FEDERER, 1998) or cubic smoothing spline (VERBYLA et al., 1999) in one dimension, may not completely account for spatial covariance. Moreover, it is extremely rare that large-scale continuous spatial variability is found only in the direction of rows or of columns, and some sort of interaction between rows and columns has to be considered in order to account for such variability (FEDERER, 1998). For such a purpose, FEDERER (1998) proposed fitting interactions between polynomials for rows and columns. However, polynomials do a poor job when fitting observations in the extremes; i.e., extreme observations have a large effect in the estimated parameters, and this is especially true for polynomials of higher degree.

Splines, as an alternative approach to polynomials, have also been used to deal with environmental heterogeneity (CAPPA and CANTET, 2007). Splines are piecewise polynomials functions from segments of lower degree polynomials (GREEN and SILVERMAN, 1994). The

positions where the segments join are called knots. Splines are able to capture most of the sinuosity present in the data with complex patterns of variation without suffering from numerical instability. A particular type of spline is the 'basic spline' (B-splines), which are local basis functions, consisting of polynomial segments of degree  $d$ , in general linear, quadratic or cubic, that has  $d - 1$  continuous derivatives at the joining points, or knots. EILERS and MARX (1996) proposed penalized splines (P-splines) in one dimension with equally spaced knots, introducing a penalty that affects first or second differences of B-spline parameters. The penalty controls the degree of smoothness while fitting the function. EILERS and MARX (2003) extended their methodology to estimate a surface in two dimensions, using the tensor product of B-splines. Either applied to one or two dimensions (EILERS and MARX, 1996; 2003), the parameters of the B-spline function are treated as fixed effects; however, the splines are closely connected to mixed models (RUPPERT et al., 2003; WAND, 2003). In a recent study, CAPPA and CANTET (2007) proposed to use tensor products of cubic B-splines based on a mixed model framework by treating the B-spline function parameters as random variables (i.e., using a covariance structure for the random knots effects) in a two-dimensional grid. They demonstrated that the methodology could account for large-scale continuous spatial variation in forest genetic evaluation of individual trials, using Bayesian techniques via Gibbs sampling, to make inferences in all dispersion parameters of the model. CAPPA et al. (unpublished) extended the method of CAPPA and CANTET (2007) demonstrating its utility in accommodating complex patterns of spatial heterogeneity in several large forest genetics trials of western hemlock (*Tsuga heterophylla* (Raf.) Sarg.) with single-tree plots design. They modeled different patterns of spatial variability having: a) small-scale variations, b) small-scale variations together with large-scale variation in one dimension (i.e., across rows or columns) and c) small-scale variations together with large-scale variation in two dimensions (i.e., across rows and columns). The new two-dimensional surface reduced the posterior means of  $\sigma_e^2$  across ten sites by 3.4 to 48.2% when compared with the 'sets in replicates' and incomplete block 'a priori' design. This resulted in an increase of the posterior mean of  $h^2$ 's from 25.0 to 76.7% and an increase in accuracy up to 3.2% for parental and offspring breeding value estimates.

Regardless of the implemented improvements in larger trials, the performance of the tensor product of B-spline bases to account for the spatial variability is unknown with limited information (i.e., fewer number of data points, in rows and columns from which the two-dimensional surface can be estimated). The goal of the present research, using two different experimental data sets, is to study the utility of surface fitting using the tensor product of B-spline to account for the spatial variation in relatively small (e.g., 576 m<sup>2</sup> and 5,705 m<sup>2</sup>) forest genetic trials, with an *a priori* simple CRD and large multiple-tree contiguous plot configuration. Additionally, we extend the CAPPA and CANTET's (2007) individual tree mixed model that used tensor product of cubic B-spline

to an individual tree mixed model with tensor product of linear, quadratic and cubic B-spline with different number of knots for rows and columns to model the spatial heterogeneity. The resulting estimates of all dispersion parameters for mixed models that include the fitted surface are finally compared with corresponding estimates from the classical CRD analysis with an individual-tree model including a plot-to-plot environmental effect.

## Material and Methods

### Western larch trial

A very small common garden trial was established at the Totem Field Research Facility, University of British Columbia, Vancouver, British Columbia (latitude 49° 15' N, longitude 123° 15' W, elevation 79m), using wind-pollinated seed samples from 15 unrelated western larch families along with an orchard "bulk" seedlot (control) provided by the British Columbia Ministry of Forests and Range's 2005 seed crop. The experimental layout was a completely randomized design, with four replications of each family and the control seedlot (N = 400 seedlings/genetic entry). The control #16 (replication 2) was the only genetic entry with missing seedlings, due to seedling shortage. Seedlings were planted in 10 x 10 contiguous square plots at a spacing of 0.3 x 0.3 m, thus the plot size was 3 by 3 m. The plots formed a rectangular grid of 11 rows and 6 columns with two unplanted plots at the corner; giving an experimental area of 33 by 18 m. Rows have coordinates  $r_i$ ,  $i = 1, 2, \dots, r = 11$  and columns coordinates  $c_j$ ,  $j = 1, 2, \dots, c = 6$ . The first plot ( $r = 1$ ,  $c = 1$ ) was set to coordinates (0, 0). During the winter of 2008 (4 years from germination), individual tree total height was measured (HT, cm). In order to fit the surface and for numerical purpose, row ( $r$ ) and column ( $c$ ) spatial coordinates of each plot were expressed in meters and all the trees of the same plot had the same  $x$  and  $y$  coordinates.

### Scots pine trial

The second and larger trial was planted in 1991 at the western part of the Czech Republic, near the city Nepomuk by the Forests of the Czech Republic, State Enterprise (latitude 49° 29' N, longitude 13° 33' E, elevation 500 m). This site is one of a series of Scots pine (*Pinus sylvestris*) trials that was initially established with the goal to predict the general combining abilities of 40 parental clones in a local seed orchard (2.24 hectares, established in 1975). The experiment consisted of half-sib progenies planted in 50-tree contiguous plots at a close spacing of 1.4 x 0.7 meters (5 x 10 plants planted at a square plot of 7 x 7 meters). On average, 200 seedlings were planted per each half-sib family (on average, 4 replicates per family). Completely randomized design was used as the layout with orchard's bulk seedlot as a control. The experiment was arranged in a rectangular scheme of 11 x 15 plots with two missing plots at one corner (plots were damaged). Rows have coordinates  $r_i$ ,  $i = 1, 2, \dots, r = 11$  and columns coordinates  $c_j$ ,  $j = 1, 2, \dots, c = 15$ . During the spring of 2008, individual tree total height was measured in centimeters (HT, cm) on all surviving trees.

Design information and descriptive statistics estimated using PROC MEANS of SAS for the western larch and Scots pine data sets are summarized in *Table 1*.

### Models of analysis

Several individual-tree additive models were evaluated for both data sets. All models, included a fixed effect of overall mean, a normally distributed random additive genetic effects ( $\mathbf{a}$ , breeding values) with covariance matrix  $\mathbf{A} \sigma_a^2$ , where  $\mathbf{A}$  is the additive relationship matrix (HENDERSON, 1984) among all trees, the additive genetic variance ( $\sigma_a^2$ ), and a normally distributed random errors ( $\mathbf{e}$ ) with mean zero and variance  $\sigma_e^2$ . The classical individual-tree model also included a normally distributed random plot effect term ( $\mathbf{p}$ ) with mean zero and variance  $\sigma_p^2$ . In the other models, and to account for the spatial variability, following CAPPA and CANTET (2007), we extended the classical model to an individual-tree mixed model with a two-dimensional surface using the tensor product of linear, quadratic and cubic B-splines. Let  $\mathbf{Y}$  be a matrix of order of rows ( $R = 110$ )  $\times$  columns ( $C = 60$  or  $75$ ) containing the tree individual observations for HT of the western larch and Scots pine trials, respectively. To transform  $\mathbf{Y}$  into a vector we used the 'vec' operator (HARVILLE, 1997; page 339), in which the  $n$  (or  $R \times C$ )  $\times$  1 vector  $\mathbf{y}$  results from stacking the columns of  $\mathbf{Y}$ :  $\mathbf{y} = \text{vec}(\mathbf{Y})$ . Then, in matrix notations each individual-tree mixed model, with a smoothed surface to account the spatial variability, can be described as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{B}\mathbf{b} + \mathbf{Z}_a\mathbf{a} + \mathbf{e} \quad [1]$$

where  $\mathbf{B}$  has dimension  $n \times (nx_r \times nx_c)$  = number of knots for rows  $\times$  number of knots for columns) and is equal to  $\mathbf{B} = (\mathbf{B}_r \otimes \mathbf{1}'_{nx_c}) \# (\mathbf{1}'_{nx_r} \otimes \mathbf{B}_c)$ ,  $\mathbf{B}_i$  ( $i = r$  or  $c$ ) are the matrices of the order  $n \times nx_i$  that contains the  $d + 1$  nonzero B-spline bases needed to express each row and column in terms of a linear, quadratic or cubic B-spline bases. Therefore, to express one row (or column) as a function of B-spline bases in  $\mathbf{B}_i$  ( $i = r$  or  $c$ ), 2 linear B-splines bases are needed, or 3 quadratic B-splines bases, or 4 cubic B-splines bases. Calculations of the  $\mathbf{B}_i$  ( $i = r$  or  $c$ ) coefficients were performed using the recursive algorithm of DE BOOR (1993). The symbols  $\otimes$  and  $\#$  indicate the Kronecker and Hadamard products of matrices, respectively (HARVILLE, 1997). The parametric vector  $\mathbf{b}$  of order  $(nx_r \times nx_c) \times 1$  contains the parameters of the tensorial product of B-splines (i.e., the random knots

effects, RKE). The distribution of the random vector  $\mathbf{b}$  is such that  $\mathbf{b} \sim N(\mathbf{0}, \mathbf{U}\sigma_b^2)$ . The scalar  $\sigma_b^2$  is the variance of the RKE for rows and columns and  $\mathbf{U}$  of order  $(nx_r \times nx_c) \times (nx_r \times nx_c)$  is the covariance structure in two dimensions for the B-spline knots. In the present study, we select the tridiagonal matrix originally proposed by GREEN and SILVERMAN (1994; page 13) and then used by DURBAN et al. (2001) to fit a fertility trend. A more detailed explanation of the two-dimensional surface ( $\mathbf{B}\mathbf{b}$ ) using the tensor product of cubic B-splines with equal number of knots for row and column, can be found in CAPPA and CANTET (2007).

The sequence of individual-tree mixed models with a smoothed surface (Model 1) fitted for the western larch and Scots pine data sets, differ in the number of knots for rows and columns and the degree of the basis functions fitted. The minimum number of knots was chosen approximately using the criterion suggested by M. WAND (see RUPPERT, 2002) who choose to place a knot every  $t$  observations, and  $t = \min(r/4$  (or  $c/4$ ), 35). Thus, up to 3 and 4 knots for row and up to 2 and 4 knots for columns were specified for the western larch and Scots pine data sets, respectively. Linear (L), quadratic (Q) and cubic (C) polynomial segments, i.e., basis functions of degree  $d = 1, 2$  and  $3$ , were considered. As in P-spline methodology, knots were chosen equally spaced across rows and columns.

### Spatial analysis of residuals

To identify spatial patterns in the both data sets, we examined the spatial distribution of the residuals for the plot means (i.e., the average of all trees from a given family plot) using a model with a fixed overall mean and random family effects. It should be noted that in this situation the resulting residuals still contain  $3/4$  of the additive genetic variance, due to the half-sibling structure of the families in both tests (excluding the bulk orchard lots). The spatial distributions of the HT residuals are illustrated in the *Figure 1a*, where the color intensity represents the magnitude of the residuals in the plot: the darker the spot, the larger the residual (note that the plot residuals are not randomly distributed in both experimental fields). Furthermore, there is clear evidence of some different residuals patterns across rows or columns, which indicate the presence of interaction between row and column position and the need for a two-dimensional smoothing.

*Table 1.* – Design information, mean with standard deviation, value minimum and maximum for the Total Height (HT, cm) in the western larch and Scots pine data sets.

|                                         | Western larch    | Scots pine        |
|-----------------------------------------|------------------|-------------------|
| Design                                  | CR               | CR                |
| Size of the trial (row $\times$ column) | 11 $\times$ 6    | 11 $\times$ 15    |
| Number total of plots                   | 64               | 163               |
| Size of plots (row $\times$ column)     | 10 $\times$ 10   | 10 $\times$ 5     |
| Size of plots in meters (m $\times$ m)  | 3 $\times$ 3     | 10 $\times$ 3.5   |
| Spacing between trees (m $\times$ m)    | 0.3 $\times$ 0.3 | 1 $\times$ 0.7    |
| Trees with records ( $n$ )              | 5,809            | 3,722             |
| Families                                | 15               | 40                |
| Individuals in the pedigree ( $q$ )     | 5,794            | 3,762             |
| Mean total height (cm) (SD)             | 72.23 (15.49)    | 967.22 (103.12)   |
| Min. – Max. total height (cm)           | 11.00 – 125.00   | 490.00 – 1,370.00 |

CR = Completely randomized; SD = Standard deviation; Min. = Minimum; Max. = Maximum.

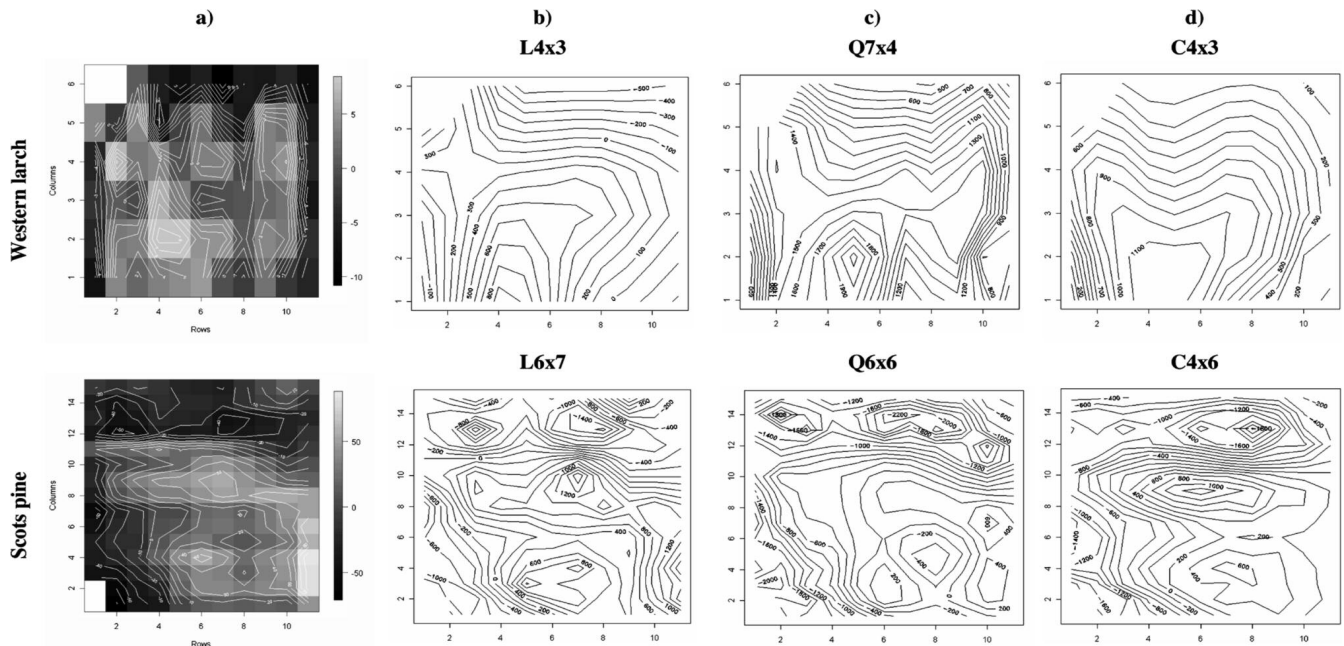


Figure 1. – First column: a) Spatial patterns of the residuals of the tree height in the western larch (top) and Scots pine (bottom) data sets. The shade intensity represents the magnitude of the residuals in the plot: the darker the spot, the larger the residual. Empty squares (upper left and lower left corner for the top and bottom plots, respectively) represent areas with no trees present. Second, third and fourth columns: b) best linear B-spline surface with  $4 \times 3$  knots (L4x3, top) and  $6 \times 7$  knots (L6x7, bottom); c) best quadratic B-spline surface with  $7 \times 4$  knots (Q7x4, top) and  $6 \times 6$  knots (Q6x6, bottom); d) best cubic B-spline surface with  $4 \times 3$  knots (C4x3, top) and  $4 \times 6$  knots (C4x6, bottom) in the western larch common garden experiment and Scots pine data set, respectively.

#### Bayesian inference and models comparison

The Bayesian approach via Gibbs sampling was used to estimate the parameters in the classical individual-tree model and in all models with a smoothed surface [1] following CAPPA and CANTET (2007). Conjugate prior densities were chosen for all parameters. To reflect a prior state of uncertainty for the fixed effects while keeping the posterior distribution proper, we selected  $\beta \sim N_p(\mathbf{0}, \mathbf{K})$  with  $\mathbf{K}$  a diagonal matrix with large elements ( $k_{ii} > 10^8$ ). For the prior distributions of  $\sigma_p^2$ ,  $\sigma_b^2$ ,  $\sigma_a^2$  and  $\sigma_e^2$ , we used a scaled inverted chi-square with hipervariances  $\delta_p^2$ ,  $\delta_b^2$ ,  $\delta_a^2$  and  $\delta_e^2$  and degrees of freedom  $v_p$ ,  $v_b$ ,  $v_a$ , and  $v_e$ , respectively. Therefore, the joint and conditional posterior densities are Gaussian for  $\beta$ ,  $\mathbf{p}$ ,  $\mathbf{b}$  and  $\mathbf{a}$  and scaled chi-square for  $\sigma_p^2$ ,  $\sigma_b^2$ ,  $\sigma_a^2$  and  $\sigma_e^2$ . At the end of each iteration, for the classical individual-tree model, the individual tree narrow-sense heritability of HT was calculated as  $h_{\text{HT}}^2 = \tilde{\sigma}_a^2 / (\tilde{\sigma}_a^2 + \tilde{\sigma}_p^2 + \tilde{\sigma}_e^2)$ , where  $\tilde{\sigma}_a^2$ ,  $\tilde{\sigma}_p^2$ ,  $\tilde{\sigma}_e^2$  are the values of the additive, plot, and error variances sampled at a given iteration. For each individual-tree mixed model with a smoothed surface (model 1),  $h_{\text{HT}}^2$  was calculated as  $h_{\text{HT}}^2 = \tilde{\sigma}_a^2 / (\tilde{\sigma}_a^2 + \tilde{\sigma}_e^2)$ .

A single Gibbs chain of 1,010,000 samples was drawn, and the first 10,000 iterates were discarded as *burn-in*. The 1,000,000 additional samples were used for computing the summary from the marginal posterior distribution. Marginal posterior densities for all parameters were estimated by the Gaussian kernel method (SILVERMAN, 1986; Chap. 2). Autocorrelations were calculated with “Bayesian Output Analysis” (BOA version 1.0.1; SMITH, 2003) for all lags from 1 to 50. Mean, mode, median, standard deviation, and 95% high posterior density interval (95% HPD), were then calculated with

BOA for all parameters from the individual marginal posteriors, under the free-software R (<http://www.r-project.org/>).

The Deviance Information Criterion (DIC; SPIEGELHALTER et al., 2002) was computed to compare the fit of each model. The DIC criterion is defined as

$$\text{DIC} = \bar{D}(\theta_M) + p_D$$

where  $\bar{D}(\theta_M)$  is the posterior mean of the deviance and  $p_D$  the “effective number of parameters”. Hence, the DIC combines a measure of model fit ( $\bar{D}(\theta_M)$ ), with a measure of model complexity ( $p_D$ ). A smaller DIC value indicates a better fit and lower degree of model complexity. Numerical details for the calculation of DIC in a multiple trait individual tree model are presented in CAPPA and CANTET (2006). Additional models comparison was provided by a visual comparison between the spatial patterns of the residuals and the resulting estimates surfaces. Finally, the accuracy of the predicted of breeding values was computed using the following expression:

$$r = \sqrt{\frac{1 - \text{PEV}}{\sigma_a^2}}$$

where PEV stands for ‘prediction error variance’ (HENDERSON, 1984) of predicted breeding values using the “Best Linear Unbiased Predictors” (BLUP’s) of parent and offspring. Spearman-rank correlations using PROC CORR of SAS were also calculated to compare whether the ranking of predicted breeding values differed between the classical individual-tree model with plot-to-plot environmental effects and the best individual-tree model with a two-dimensional surface.

Computer programs employed to carry out the Bayesian inference, as well as to solve the mixed models equations and to obtain corresponding accuracies of all models analyses were developed in FORTRAN (the FORTRAN program is available upon request).

## Results

The values of DIC and marginal posterior means with the respective 95 % high posterior density interval of  $\sigma_a^2$ ,  $\sigma_p^2$ ,  $\sigma_b^2$ ,  $\sigma_e^2$  and  $h^2_{HT}$  for the mixed models that include the fitted surface, together with corresponding estimates for the classical individual-tree model (that included plots effects), are summarized for the two data sets in *Tables 2* and *3*. Note that models accounting for site variability with a two-dimensional surface had a smaller DIC (i.e., better fit) than the classical RCD model that included random plot environmental effects, with the exception of the model Q3x3, and all models that fit a cubic two-dimensional B-spline in the western larch experiment (*Table 2*). This was likely due to the higher degree (cubic) of the B-spline and the relatively small amount of spatial data points for columns (i.e., only 6 data points) for estimating the two-dimensional surface. These results are consistent with those obtained by CAPPA and CANTET (2007), who observed lower values of DIC for models including tensor products of cubic B-splines, as compared to the DIC of a model with an *a priori* design, in a much larger progeny test. In the western larch experiment, model Q7x4 showed the smallest DIC, followed by the model L4x3. In the Scots pine trial, model L6x7 showed the smallest DIC, followed by models L6x6 and Q6x6 (*Table 3*). In general, spatial models with larger fit order required the lowest number of knots for row and column for the best

fits (i.e., smaller DIC), so there was an offsetting between the order of fit and the number of knots. For example, in the Scots pine data set, the best model for the linear B-splines utilized 6 knots for row and 7 for column (L6x7), whereas the best model in the cubic B-spline was the model with 4 knots for row and 6 for column (C4x6) (*Table 3*). The presence of spatial effects could be observed in *Figure 1b-d*, which displays the estimated smoothest surface for the mixed model with best fit (i.e., smaller DIC) for each of the three different orders of the B-spline fitted in the western larch experiment (*Figure 1b-d*, first row) and in the Scots pine trial (*Figure 1b-d*, second row). When we compared the smoothest surface estimated for the best model in each order of fit (*Figure 1b-d*), models Q7x4 and L6x7 showed the highest degree of smoothing for the western larch and Scots pine data sets, respectively. Models Q7x4 and L6x7 were the best in capturing the spatial variability of the two data sets, observed in the respective residual spatial pattern plots (*Figure 1a*). That is, visual comparison between the estimated smoother surfaces of models Q7x4 and L6x7 (*Figure 1c-top* and *Figure 1b-bottom*, respectively) and the corresponding spatial patterns of the residuals (*Figure 1a*) highlights only slight differences.

The fit of smoothed surfaces produced a reduction of the posterior means of  $\sigma_e^2$  in both data sets, again except for the model Q3x3 and all the models utilizing a cubic two-dimensional B-spline in the western larch experiment (*Tables 2* and *3*). In addition, increasing the number of knots for row and column tended to reduce the variance error estimates. The  $\sigma_e^2$  estimate in the western larch trial for the classical RCD model was 150.28, which was, respectively, only 3.99 and 3.62% higher than those estimates from the best models for the qua-

*Table 2.* – Deviance Information Criterion (DIC) and posterior means (95% high posterior density interval) for the additive genetic variances ( $\sigma_a^2$ ), the plot variance ( $\sigma_p^2$ ), the variance of the RKE ( $\sigma_b^2$ ), the error variance ( $\sigma_e^2$ ), and the individual narrow-sense heritability of height ( $h^2_{HT}$ ) in the western larch common garden experiment. Models with smaller DIC (i.e., better fits) in each order of B-splines are given in bold type.

| Model <sup>a</sup> | DIC             | $\sigma_a^2$               | $\sigma_p^2$       | $\sigma_b^2$                   | $\sigma_e^2$                 | $h^2_{HT}$             |
|--------------------|-----------------|----------------------------|--------------------|--------------------------------|------------------------------|------------------------|
| <b>Classical</b>   | 22781.48        | 63.11(34.98-102.99)        | 24.50(17.30-33.78) | -                              | 150.28(120.00-172.37)        | 0.26(0.15-0.42)        |
| <b>L3x2</b>        | 22734.46        | 78.88(48.56-120.35)        | -                  | 209.53(108.41-376.58)          | 146.85(115.40-170.57)        | 0.35(0.22-0.51)        |
| <b>L3x3</b>        | 22753.29        | 74.36(45.40-114.17)        | -                  | 168.68(93.12-288.14)           | 148.21(117.96-170.95)        | 0.33(0.21-0.49)        |
| <b>L4x3</b>        | <b>22707.50</b> | <b>76.51(47.24-117.45)</b> | -                  | <b>171.42(98.42-283.36)</b>    | <b>144.84(113.69-167.73)</b> | <b>0.34(0.22-0.51)</b> |
| <b>L5x3</b>        | 22716.37        | 76.13(47.01-115.78)        | -                  | 141.89(84.98-226.39)           | 145.20(115.14-167.99)        | 0.34(0.22-0.50)        |
| <b>Q3x3</b>        | 22788.11        | 69.39(42.07-107.36)        | -                  | 611.22(319.88-1068.31)         | 150.52(121.58-172.07)        | 0.31(0.20-0.47)        |
| <b>Q4x3</b>        | 22751.21        | 72.12(44.32-111.65)        | -                  | 409.66(225.25693.55)           | 147.40(117.27-169.29)        | 0.33(0.21-0.49)        |
| <b>Q5x3</b>        | 22725.30        | 71.68(43.98-110.29)        | -                  | 373.23(215.23-608.08)          | 146.54(117.24-168.26)        | 0.33(0.21-0.48)        |
| <b>Q7x4</b>        | <b>22687.85</b> | <b>73.74(45.03-113.07)</b> | -                  | <b>322.30(196.74-497.99)</b>   | <b>144.29(114.40-166.78)</b> | <b>0.34(0.21-0.50)</b> |
| <b>Q8x4</b>        | 22758.46        | 67.54(40.15-105.61)        | -                  | 313.16(191.50-481.77)          | 148.53(119.57-170.05)        | 0.31(0.19-0.47)        |
| <b>C3x2</b>        | 22924.71        | 63.98(38.60-100.09)        | -                  | 828.61(316.05-1738.73)         | 158.99(131.41-179.26)        | 0.29(0.18-0.43)        |
| <b>C3x3</b>        | 22869.63        | 64.52(39.50-102.42)        | -                  | 5285.18(257.11-15086.46)       | 155.75(127.47-176.61)        | 0.29(0.18-0.45)        |
| <b>C4x3</b>        | <b>22798.79</b> | <b>70.98(43.56-110.39)</b> | -                  | <b>1147.26(451.21-2365.84)</b> | <b>150.34(120.26-172.01)</b> | <b>0.32(0.20-0.48)</b> |
| <b>C5x3</b>        | 22802.13        | 69.40(42.20-107.37)        | -                  | 490.59(237.09-918.82)          | 151.20(122.30-172.63)        | 0.31(0.20-0.47)        |

**NOTE:** <sup>a</sup> Model Classical: non-spatial individual-tree model including a plot-to-plot environmental effect.

Model  $Lnx_r \times nx_c$ : individual-tree mixed model including a linear (L) B-splines with  $nx_r$ , number of knots for row and  $nx_c$ , number of knots for column.

Model  $Qnx_r \times nx_c$ : individual-tree mixed model including a quadratic (Q) B-splines with  $nx_r$ , number of knots for row and  $nx_c$ , number of knots for column.

Model  $Cnx_r \times nx_c$ : individual-tree mixed model including a cubic (C) B-splines with  $nx_r$ , number of knots for row and  $nx_c$ , number of knots for column.

Table 3. – Deviance Information Criterion (DIC) and posterior means (95% high posterior density interval) for the additive genetic variances ( $\sigma_A^2$ ), the plot variance ( $\sigma_p^2$ ), the variance of the RKE ( $\sigma_b^2$ ), the error variance ( $\sigma_e^2$ ), and the individual narrow-sense heritability of height ( $h_{HT}^2$ ) in the Scots pine data set. Models with smaller DIC (i.e., better fits) in each order of B-splines are given in bold type.

| Model <sup>a</sup> | DIC             | $\sigma_A^2$                    | $\sigma_p^2$             | $\sigma_b^2$                       | $\sigma_e^2$                    | $h_{HT}^2$             |
|--------------------|-----------------|---------------------------------|--------------------------|------------------------------------|---------------------------------|------------------------|
| <b>Classical</b>   | 20907.68        | 1662.48(892.44-2720.13)         | 1930.97(1518.34-2412.14) | -                                  | 7149.94(6305.95-7821.87)        | 0.15(0.08-0.25)        |
| <b>L4x4</b>        | 20740.27        | 3722.05(2523.61-5262.07)        | -                        | 7363.57(4433.08-11696.42)          | 6203.17(5006.40-7178.41)        | 0.37(0.26-0.51)        |
| <b>L4x5</b>        | 20774.33        | 3548.96(2389.22-5063.70)        | -                        | 6675.78(4151.55-10292.52)          | 6300.73(5123.39-7249.82)        | 0.36(0.25-0.50)        |
| <b>L4x6</b>        | 20759.19        | 3382.83(2277.86-4812.13)        | -                        | 7421.52(4715.71-11246.13)          | 6265.16(5148.04-7173.14)        | 0.35(0.24-0.48)        |
| <b>L5x5</b>        | 20724.40        | 3630.23(2474.97-5084.28)        | -                        | 7041.15(4470.43-10646.13)          | 6152.30(5017.92-7092.85)        | 0.37(0.26-0.50)        |
| <b>L5x6</b>        | 20710.66        | 3489.58(2385.98-4879.26)        | -                        | 7505.81(4905.69-11074.49)          | 6105.85(5021.85-7005.52)        | 0.36(0.26-0.50)        |
| <b>L6x6</b>        | 20699.19        | 3496.53(2341.02-5019.65)        | -                        | 6907.24(4633.71-9967.86)           | 6077.32(4896.07-7016.87)        | 0.36(0.25-0.50)        |
| <b>L6x7</b>        | <b>20681.74</b> | <b>3582.63(2428.36-5071.18)</b> | -                        | <b>6361.22(4341.07-9036.77)</b>    | <b>6008.20(4851.89-6947.07)</b> | <b>0.37(0.26-0.51)</b> |
| <b>L7x7</b>        | 20714.90        | 3365.00(2257.84-4803.66)        | -                        | 6628.76(4562.96-9344.81)           | 6138.55(5021.99-7041.89)        | 0.35(0.24-0.49)        |
| <b>Q4x4</b>        | 20770.97        | 3631.69(2463.06-5124.91)        | -                        | 11764.49(6545.524-19656.74)        | 6295.32(5132.22-7248.04)        | 0.36(0.26-0.50)        |
| <b>Q4x5</b>        | 20720.85        | 3625.28(2445.23-5151.75)        | -                        | 18522.47(10407.33-30446.55)        | 6126.49(4942.12-7086.85)        | 0.37(0.26-0.51)        |
| <b>Q4x6</b>        | 20729.44        | 3624.03(2430.57-5173.25)        | -                        | 14853.05(8533.39-23978.02)         | 6157.79(4954.47-7130.22)        | 0.37(0.26-0.51)        |
| <b>Q5x5</b>        | 20753.97        | 3426.22(2327.22-4826.44)        | -                        | 17702.99(10282.48-28282.46)        | 6248.73(5153.27-7148.08)        | 0.35(0.25-0.48)        |
| <b>Q5x6</b>        | 20757.21        | 3453.44(2366.54-4845.98)        | -                        | 14215.15(8417.32-22425.73)         | 6251.83(5169.03-7141.72)        | 0.36(0.25-0.48)        |
| <b>Q6x6</b>        | <b>20708.02</b> | <b>3610.92(2429.62-5153.26)</b> | -                        | <b>13252.47(8007.95-20587.21)</b>  | <b>6096.27(4902.61-7057.89)</b> | <b>0.37(0.26-0.52)</b> |
| <b>Q6x7</b>        | 20733.71        | 3445.58(2340.18-4874.47)        | -                        | 11124.09(7078.33-16662.23)         | 6179.20(5065.38-7087.84)        | 0.36(0.25-0.49)        |
| <b>C4x4</b>        | 20728.96        | 4125.24(2821.48-5759.27)        | -                        | 11197.58(6047.55-19362.34)         | 6163.41(4902.65-7214.26)        | 0.40(0.28-0.54)        |
| <b>C4x5</b>        | 20745.30        | 3873.73(2624.82-5482.45)        | -                        | 21055.76(9262.5845-42809.48)       | 6209.64(4969.07-7220.29)        | 0.38(0.27-0.52)        |
| <b>C4x6</b>        | <b>20715.81</b> | <b>3713.53(2501.95-5273.65)</b> | -                        | <b>22849.99(12725.52-38244.31)</b> | <b>6118.42(4909.31-7099.33)</b> | <b>0.38(0.26-0.52)</b> |
| <b>C5x5</b>        | 20785.91        | 3690.63(2504.01-5192.61)        | -                        | 17970.13(8542.88-34343.06)         | 6350.13(5179.42-7314.23)        | 0.37(0.26-0.50)        |

NOTE: <sup>a</sup> Model Classical: non-spatial individual-tree model including a plot-to-plot environmental effect.

Model  $L_{n_x \times n_x}$ : individual-tree mixed model including a linear (L) B-splines with  $n_x$ , number of knots for row and  $n_x$ , number of knots for column.

Model  $L_{n_x \times n_x}$ : individual-tree mixed model including a quadratic (Q) B-splines with  $n_x$ , number of knots for row and  $n_x$ , number of knots for column.

Model  $L_{n_x \times n_x}$ : individual-tree mixed model including a cubic (C) B-splines with  $n_x$ , number of knots for row and  $n_x$ , number of knots for column.

dratic (Q7x4) and linear (L4x3) B-splines (Table 2). For the Scots pine site,  $\sigma_e^2$  were substantially reduced to 15.97, 14.74 and 14.43% with fitting the best linear (L6x7), quadratic (Q6x6), and cubic (C4x6) B-splines, respectively, over the classic individual-tree model. Clearly, these reductions of the error variances in both data sets confirm that there are spatial variations larger than the plot sizes that are not adequately accounted for by a plot-to-plot environmental variance term. Conversely, smoothing surfaces fit in both data sets resulted in higher posterior means of  $\sigma_a^2$  for all models (Tables 2 and 3). The estimated posterior means of  $h_{HT}^2$  for the models that fitted a two-dimensional surface for the western larch (from 0.29 to 0.35) and Scots pine (from 0.35 to 0.40) data sets were consistently higher than those estimated from the classic model (0.26 and 0.15, respectively). As expected, the estimated posterior means for  $\sigma_b^2$  increased from the linear to cubic order of fit of the B-splines and tended to decrease as the number of knots increased in both data sets. The 95% HPD for  $\sigma_a^2$ ,  $\sigma_p^2$ ,  $\sigma_b^2$ ,  $\sigma_e^2$  and  $h_{HT}^2$  were shifted away from the zero value for all parameters.

The average prediction accuracy of breeding values, calculated from model Q7x4 (the one with the smallest DIC) in western larch, was higher for parents (0.94) and progeny (0.66), than corresponding values (0.81 and 0.59) calculated from the classical CRD model. In Scots pine, the average accuracy of prediction of breeding values, calculated from model L6x7 (the smallest DIC), was

also higher for parents (0.92) and progeny (0.68), than their corresponding values (0.63 and 0.47) calculated from the CRD classical model. Large improvements in prediction accuracy for parents (16.05 versus 46.03%) and offspring (11.86 versus 44.68%) in the Scots pine over the western larch are also in line with the larger increment of the values of the  $h_{HT}^2$ . The Spearman-rank correlation between predicted breeding values from the best spatial model in the western larch (Q7x4) and Scots pine (L6x7) and the classical RCD model were 0.89 and 0.92 for parents and 0.92 and 0.94 for offspring, respectively, indicating that there was some rank change among parents and progenies in both data sets.

## Discussion

Forest genetics trials of typically small size (< 1 ha) with close plantings and under intensive site management can provide the illusion of a corresponding small environmental variability. In that instance, any simple type of *a priori* experimental field design based on randomization theory should account for any of the minor site variability (SCHABENBERGER and GOTWAY, 2005; page 301). However, most published results indicate that in these types of trials a large component of the spatial variation still resides in the error term (e.g., EL-KASSABY and PARK, 1993; REHFELDT, 1995; KRAKOWSKI et al., 2005; ST. CLAIR, 2006). Thus, the application of appropriate *a posteriori* statistical techniques for removing the “unac-

counted for” spatial variability that cannot be dealt with through randomization of the experimental unit (i.e., plots) is necessary. Here we applied B-splines in a two-dimensional surface, using tensor product of linear, quadratic and cubic B-spline bases with different and equal number of knots for rows and columns in an individual-tree mixed model, to remove the environmental spatial variability larger than 9 m<sup>2</sup> in the relatively small 576 m<sup>2</sup> western larch common garden and larger than 35 m<sup>2</sup> in the 5,705 m<sup>2</sup> Scots pine progeny trial. Our results showed that the best model which included a tensor product of B-splines reduced the posterior means of the error variance ( $\sigma_e^2$ ) in the two experiments by 3.99 and 15.97%. The fit of a smoothed surface to both data sets, instead of the *a priori* RCD, consistently increased the estimated posterior mean of the additive genetic variance ( $\sigma_a^2$ , Tables 2 and 3). In the Scots pine data set the increase in  $\sigma_a^2$  approximately doubled, over the non-spatial analysis (as were the heritabilities); however, the change in the  $\sigma_a^2$  using different *a posteriori* spatial approaches showed diverse results. JOYCE et al. (2002) reported a consistent decrease in the  $\sigma_a^2$  in a farm-field test of black spruce, using a two-step approach: first, they examined the large-scale variation with median polishing methods, followed by detecting small-scale variations with a nearest neighbour technique. Including a spatial correlated error, increases or decreases in the values of  $\sigma_a^2$  were reported when compared with different estimates of the *a priori* experimental (field) design (YE and JAYAWICKRAMA, 2008; COSTA e SILVA, 2001; DUTKOWSKI et al., 2002, 2006; KUSNANDAR and GALWEY, 2000; ANEKONDA and LIBBY, 1996). CAPPA et al. (unpublished), using the method applied in this present study, showed small but also inconsistent effect on the  $\sigma_a^2$  from 10 large progeny trials of western hemlock with *a priori* “sets in replicates” and an incomplete block design with single-tree plots configuration. However, MAGNUSSEN (1993), using simulated data, showed that the estimate  $\sigma_a^2$  may be overestimated, even with modified nearest-neighbour spatial adjustments, in forest genetic trials with multiple-tree and contiguous plot configuration.

In the present study, large square and rectangular multiple-tree and contiguous plots of 10x10 (9 m<sup>2</sup>) and 10x5 (35 m<sup>2</sup>) trees made it difficult to definitively separate the genetic and environmental effects within experimental units (i.e., plots). In multiple-tree and contiguous plots designs, substantial environmental covariance among member of a family is confounded with genetic covariance in a given plot (LOO-DINKINS, 1992). The degree of confounding depends on the size of the plots and the patterns of environment variability. In general, the larger the plot, the more difficult it is to cleanly separate genetic from environmental effects. On the other hand, single-tree plots maximize the variety of the environmental variations sampled more than multiple-tree contiguous plots (e.g. LOO-DINKINS and TAUER, 1987; GEZAN et al., 2006). Therefore, the fit of any model with spatial components in relatively small forest genetic trials, with large multiple-tree contiguous plots design where all trees within each plot have the same *x* and *y* coordinates, will tend to be less efficient to account for the spatial variation as compared with single-tree plots configuration. Single-tree plots maximize the number of

data points from which the two-dimensional surface will be estimated (as every tree contributes its own *x* and *y* coordinates), and should result in the most accurate modeling of spatial variability.

The average accuracy of breeding values from parents and offspring estimated with the best spatial model were higher than the corresponding values estimated with the classical RCD due to the increase in the estimated additive variance and decrease of estimated error variance (Tables 2 and 3). The larger increases in heritabilities in the Scots pine trial were associated with larger increase in the estimated additive variance and larger decrease of estimated error variance, relative to that of western larch. These results are in agreement with several studies that compared spatial models with different *a priori* designs (COSTA e SILVA et al., 2001; DUTKOWSKI et al., 2002, 2006; ZAS, 2006; YE and JAYAWICKRAMA, 2008). For example, compared to the randomized complete block designs, increases in accuracy from spatial models were reported by COSTA e SILVA et al. (2001) for tree height and ZAS (2006) for tree diameter. COSTA e SILVA et al. (2001) analyzed twelve trials and found up to 71% increases in accuracy of predicted additive effects of the parents and offspring. Similarly, ZAS (2006) reported substantial increases in the accuracy of BLUPs of family effects, ranging from 0.40–0.63 up to 0.72–0.79 after correcting for spatially correlated variation. YE and JAYAWICKRAMA (2008) found that the accuracy of breeding value prediction increased by up to 20% in comparison with the “sets in replicates” and “replicates in sets” *a priori* designs in 275 large Douglas-fir (*Pseudotsuga menziesii* (Mirb.) Franco) progeny trials. A smaller gain in accuracy was found by DUTKOWSKI et al. (2002; 2006), but still in the direction of the spatial model being superior to the *a priori* traditional design-based model. A substantial fraction of the gain in accuracy is due to the fact that not all environmental spatial variability (i.e., larger than 9 m<sup>2</sup> in the 576 m<sup>2</sup> western larch common garden and larger than 35 m<sup>2</sup> in the 5,705 m<sup>2</sup> in Scots pine progeny trial) is accounted for in the plot-to-plot variability; i.e., variation that otherwise would go to the error variance. Therefore, analysis of data displaying spatial variations larger to plot size with spatial models will most likely improve the accuracy of selection as compared with an analysis using the CRD model.

Following CAPPA and CANTET (2007), we used B-splines with equally spaced knots to account for the spatial variation in forest genetic trials. In this approach, the crucial parameter is the penalty or smoothing factor  $\lambda$  (equation [5] in CAPPA and CANTET, 2007), and the number of knots in the spline is not as critical to the fit as long as they are “sufficiently” present (EILERS and MARX, 1996). In the mixed model approach to P-splines,  $\lambda$  is the ratio  $\sigma_e^2 / \sigma_b^2$  (CAPPA and CANTET, 2007). Looking at Tables 2 and 3, one may infer that the magnitude of  $\sigma_b^2$  (the denominator of  $\lambda$ ) was sensitive to the number of knots, as compared to the other variance components, tend to decrease when the number of knots fitted increased. It is known that the fit of very few knots produces bias, which rapidly decreases as the number of knots increases (RUPPERT, 2002). In other words, choosing a small number of knots would lead to a smoother

surface that is not flexible enough to capture the spatial variability in the data. Once the minimum number is reached, increasing the number of knots often gives satisfactory fits (RUPPERT, 2002), as long as the general pattern of environmental variation is adequately addressed. For example, CAPP and CANTET (2007) showed that the change in the DIC and the estimated genetic parameters between the models with 12x12 knots (a better fit) and 18x18 knots for row and column were minimal. Similarly, in the present study, once the best number of knots was reached, increasing the number of knots increased the DIC (i.e., a worse fit). However, in all cases this change was relatively minor. For example, when we fitted a two-dimensional quadratic B-splines in the western larch experiment, the model with 7x4 knots (model Q7x4) displayed the lowest DIC value (22,687.85), the addition of one knot in the row (model Q8x4) increased the DIC value (to 22,758.46) and decreased the heritability from 0.34 to 0.31. When we fitted a two-dimensional linear B-splines in the Scots pine trial, increasing the number of knots in row from 6 to 7 (i.e., models L6x7 vs. L7x7) also increased the DIC value (i.e., DIC values of 20,681.74 vs. 20,714.90, respectively) and decreased the  $h^2$  estimates (0.37 and 0.35, respectively). Moreover, with only a few number of data points for the rows and columns definitions for the western larch (i.e., 11 rows and 6 columns) and for the Scots pine (i.e., 11 rows and 15 columns), a large number of knots results in an excessively “wiggly” estimates surfaces, tending to overfit the data (data not shown).

The number of knots chosen affects the amount of smoothing applied to the data by controlling the number of piecewise fits. Additionally, the amount of smoothing of the surface between knots also depends on the degree of the individual polynomial segments (EILERS and MARX, 1996). The polynomial segments are commonly of degrees less or equal to three: i.e., linear, quadratic or cubic. The cubic B-spline bases are used instead of linear or quadratic bases, to allow for more flexibility in fitting peaks and valleys (i.e., to model flexible surfaces) of complex and different patterns of spatial variability, when a large number of data points for rows and columns are available (CAPP and CANTET, 2007; CAPP et al. (unpublished)). In the present study, we extended the cubic two-dimensional B-spline proposed by CAPP and CANTET (2007) to linear, quadratic, and cubic B-splines with different number of knots for row and column, but due to the relatively small sizes of the tests we examined, the cubic bases model did not yield the lowest DIC. However, there are several examples of the use of spline functions with different degree only in one dimension when analyzing breeding data. Thus, animal breeders have used linear splines to model growth in beef cattle (BOHMANOVA et al., 2005; IWAISAKI et al., 2005) or cubic B-spline to smoothing of lactation curves (WHITE et al., 1999) and to model the effects of management unit and time (CANTET et al., 2005). MEYER (2005) used linear to cubic B-splines to model growth of Australian Angus cattle. In forest tree breeding, CORNILLON et al. (2003) used quadratic B-splines to model time functional breeding values of clones in *Eucalyptus* using a fixed effects model. MAGNUSSEN and YANCHUK (1994) fitted spline functions to observed data so as to estimate

the individual heights at non-recorded times from Douglas-fir trees. The resulting data were then used to predict breeding values at non-recorded ages and genetic dispersion parameters.

In general, if the number of knots for row and column is large with respect to the number of available data points, linear B-spline bases may yield a reasonable description of the environmental variability. In the Scots pine data, the highest DIC values (i.e., the worst fits) for a linear B-spline with a few knots for row (i.e., 4) and less than 6 knots for columns, suggested that spatial variability might not have been modeled adequately by linear polynomials segments compared with quadratic B-spline. On the other hand, if the number of knots for row is up to 4 it may well be that those polynomial segments between the knots of degree 1, are acceptable and then a linear B-spline bases will suffice. Meanwhile, cubic B-spline with up to 4 knots for row and column introduces ‘more flexibility’ than needed and resulted in higher values of DIC than linear and quadratic B-splines. In the western larch data set, all the models with linear B-spline yielded better fit (i.e., smaller DIC) than quadratic and cubic B-spline, which again suggests that linear B-splines are sufficient to model reasonably the spatial variability, when a relatively small amount of information available (i.e., fewest data points).

## Conclusions

The tensor product of basis functions of B-spline for rows and columns does provide a useful new alternative to model patterns of spatial variability *a posteriori* even in small forest genetics trials. It was found to be well suited to adjust for spatial variation for tree height data than the simple completely randomized design that attempts to reduce the spatial heterogeneity through randomization of experimental units (i.e., plots). However, the results varied between the two data sets we examined. This is expected as improvements to the model fit are determined by several factors, including the size of the trials, the original *a priori* design in place, and the underlying spatial variation present in the test sites. This approach will be most effective when a single-tree plots design is employed, in order that maximizes the number of spatial data points from which the two-dimensional surface will be estimated. However, there are still many forest genetics trials with completely randomized design and large multiple-tree and contiguous plots configurations that must be evaluated with plausible accuracy. In relatively small forest genetic trials (<1 ha) with large multiple-tree plots design, this methodology should be used particularly when there is strong evidence of spatial correlations between plots, as present in the two data sets examined here. Further research to compare our methods with other spatial techniques by computer simulation would certainly be interesting, but this task is beyond the scope of the present study.

## Acknowledgements

Funds from British Columbia Ministry of Forests and Range to E.P.C. and A.D.Y., the Johnson’s Family Forest

Biotechnology Endowment, the Natural Sciences and Engineering Research Council of Canada- Discovery and the IRC grants to Y.A.E., and the Czech Science Foundation (GACR, grant 521/07/9337) and the National Agency for Agricultural Research (NAZV, grant QH81172) to M.L. are greatly appreciated. We thank T. FUNDA, I. FUNDOVA, B. LAI, J. KANAK, and J. KLAPSTE for technical support.

## References

- ANEKONDA, T. S. and W. J. LIBBY (1996): Effectiveness of nearestneighbor data adjustment in a clonal test of Redwood. *Silvae Genet.* **45**(1): 46–51.
- BOHMANOVA, J., I. MISZTAL and J. K. BERTRAND (2005): Studies on multiple trait and random regression models for genetic evaluation of beef cattle for growth. *J Anim Sci* **83**: 62–67.
- CANTET, R. J. C., A. N. BIRCHMEIER, A. W. CANAZA CAYO and C. FIORETT (2005): Semiparametric animal models via penalized splines as alternatives to models with contemporary groups. *J Anim Sci* **83**: 2482–2494.
- CAPPA, E. P. and R. J. C. CANTET (2006): Bayesian inference for normal multiple-trait individual-tree models with missing records via full conjugate Gibbs. *Can J For Res* **36**: 1276–1285.
- CAPPA, E. P. and R. J. C. CANTET (2007): Bayesian estimation of a surface to account for a spatial trend using penalized splines in an individual-tree mixed model. *Can J For Res* **37**: 2677–2688.
- COSTA e SILVA, J., G. W. DUTKOWSKI and A. R. GILMOUR (2001): Analysis of early tree height in forest genetic trials is enhanced by including a spatially correlated residual. *Can J For Res* **31**: 1887–1893.
- CORNILLON, P. A., L. SAINT-ANDRE, J. M. BOUVET, P. VIGNERON, A. SAYA and R. GOUMA (2003): Using B-splines for growth curve classification: applications to selection of eucalypt clones. *Forest Ecology and Management* **176**: 75–85.
- DE BOOR, C. (1993): B(asic)-spline basics. *Fundamental Developments of Computer-Aided Geometric Modeling*. Edited by L. PIEGL, Academic Press, San Diego, CA.
- DURBAN, M., I. CURRIE and R. KEMPTON (2001): Adjusting for fertility and competition in variety trials. *J Agric Sci (Camb)* **136**: 129–140.
- DUTKOWSKI, G. W., J. COSTA e SILVA, A. R. GILMOUR and G. A. LOPEZ (2002): Spatial analysis methods for forest genetic trials. *Can J For Res* **32**: 2201–2214.
- DUTKOWSKI, G. W., J. COSTA e SILVA, A. R. GILMOUR, H. WELLENDOFF and A. AGUIAR (2006): Spatial analysis enhances modeling of a wide variety of traits in forest genetic trials. *Can J For Res* **36**: 1851–1870.
- EILERS, P. H. C. and B. D. MARX (1996): Flexible smoothing with B-splines and penalties (with comments and rejoinder). *Stat Sci* **11**: 89–121.
- EILERS, P. H. C. and B. D. MARX (2003): Multivariate calibration with temperature interaction using two-dimensional penalized signal regression. *Chemometr. Intell Lab Syst* **66**: 159–174.
- EL-KASSABY, Y. A. and Y. S. PARK (1993): Genetic variation and correlation in growth, biomass traits, and vegetative phenology of a 3-year-old Douglas-fir common garden at different spacings. *Silvae Genet* **42**: 289–297.
- ERICSSON, T. (1997): Enhanced heritabilities and best linear unbiased predictors through appropriate blocking of progeny trials. *Can J For Res* **27**: 2097–2101.
- FEDERER, W. T. (1998): Recovery of interblock, intergradient, and intervarietal information in incomplete block and lattice rectangle designed experiments. *Biometrics* **54**: 471–481.
- FINLEY, A. O., S. BANERJEE, P. WALDMANN and T. ERICSSON (2009): Hierarchical spatial modeling of additive and dominance genetic variance for large spatial trial data sets. *Biometrics* **65**: 441–451.
- GILMOUR, A. R., B. R. CULLIS and A. P. VERBYLA (1997): Accounting for natural and extraneous variation in the analysis of field experiments. *J Agric Biol Environ Stat* **2**: 269–293.
- GEZAN, S. A., D. A. HUBER and T. L. WHITE (2006): Post hoc blocking to improve heritability and precision of best linear unbiased genetic predictions. *Can J For Res* **36**: 2141–2147.
- GREEN, P. J. and B. W. SILVERMAN (1994): *Nonparametric Regression and Generalized Linear Model*. Chapman & Hall, London, UK.
- GRONDONA, M. O., J. CROSSA, P. N. FOX and W. H. PFEIFFER (1996): Analysis of variety yield trials using two-dimensional separable ARIMA processes. *Biometrics* **52**: 763–770.
- HENDERSON, C. R. (1984): *Applications of Linear Models in Animal Breeding*. Canada, University of Guelph, Guelph, Ont.
- HAMANN, A., M. KOSHY and G. NAMKOONG (2002): Improving precision of breeding values by removing spatially autocorrelated variation in forestry field experiments. *Silvae Genet* **51**: 210–215.
- HARVILLE, D. A. (1997): *Matrix algebra from a statistician's perspective*. Springer-Verlag, New York.
- IWAISAKI, H., S. TSURUTA, I. MISZTAL and J. K. BERTRAND (2005): Genetic parameters estimated with multi-trait and linear spline-random regression models using Gelbvieh early growth data. *J Anim Sci* **83**: 499–506.
- JOYCE, D., R. FORD and Y. B. FU (2002): Spatial patterns of tree height variations in a black spruce farm-field progeny test and neighbors-adjusted estimations of genetic parameters. *Silvae Genet* **51**: 13–18.
- KRAKOWSKI, J., Y. S. PARK and Y. A. EL-KASSABY (2005): Early testing of Douglas-fir: wood density and ring width. *For Genet* **12**: 99–105.
- KROON, J., B. ANDERSSON and T. J. MULLIN (2008): Genetic variation in the diameter-height relationship in Scots pine (*Pinus sylvestris*). *Can J For Res* **38**: 1493–1503.
- KUSNANDAR, D. and N. GALWEY (2000): A Proposed Method for Estimation of Genetic Parameters on Forest Trees Without Raising Progeny: Critical Evaluation and Refinement. *Silvae Genet* **49**: 15–21.
- LOO-DINKINS, J. A. and C. G. TAUER (1987): Statistical efficiency of six progeny test field designs on three loblolly pine (*Pinus taeda* L.) site types. *Can J For Res* **17**: 1066–1070.
- LOO-DINKINS, J. (1992): Field test design. *In: Handbook of quantitative forest genetics*. Edited by L. FINS, S. FRIEDMAN, and J. V. BROTSCHOL. Kluwer Academic Publishers, Dordrecht, the Netherlands. pp. 96–139.
- LOPEZ, G. A., B. M. POTTS, G. W. DUTKOWSKI, L. A. APIOLAZA and P. GELID (2002): Genetic variation and inter-trait correlations in *Eucalyptus globulus* base population trials in Argentina. *Forest Genetics* **9**: 223–237.
- MAGNUSSEN, S. (1990): Application and comparison of spatial models in analyzing tree-genetics field trials. *Can J For Res* **20**: 536–546.
- MAGNUSSEN, S. (1993): Bias in genetic variance estimates due to spatial autocorrelation. *Theor Appl Genet* **86**: 349–355.

- MAGNUSSEN, S. and A. D. YANCHUK (1994): Time trends of predicted breeding values in selected crosses of coastal Douglas-fir in British Columbia: a methodological study. *For Sci* **40**: 663–685.
- MEYER, K. (2005): Random regression analyses using B-splines to model growth of Australian Angus cattle. *Genet Sel Evol* **37**: 473–500.
- REHFELDT, G. E. (1995): Genetic variation, climate models and the ecological genetics of *Larix occidentalis*. *For Ecol Manage* **78**: 21–37.
- RUPPERT, D. (2002): Selecting the number of knots for penalized splines. *Journal of Computational and Graphical Statistics* **11**: 735–757.
- RUPPERT, D., M. P. WAND and R. J. CARROLL (2003): *Semi-parametric Regression*. Cambridge Univ Press, Cambridge, UK.
- SAENZ-ROMERO, C., E. V. NORDHEIM, R. P. GURIES and P. M. CRUMP (2001): A case study of a provenance/progeny test using trend analysis with correlated errors and SAS PROC MIXED. *Silvae Genet* **50**: 127–135.
- SCHABENBERGER, O. and C. A. GOTWAY (2005): *Statistical Methods for Spatial Data Analysis*. Boca Raton: Chapman & Hall.
- SILVERMAN, B. (1986): *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London.
- SMITH, B. J. (2003): *Bayesian Output Analysis Program (BOA) version 1.0 user's manual*. <http://www.public-health.uiowa.edu/boa/Home.html>. Cited 14 Aug 2008.
- SPIEGELHALTER, D. J., N. G. BEST, B. P. CARLIN and A. VAN DER LINDE (2002): Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society Series B* **64**: 583–639.
- ST. CLAIR, J. B. (2006): Genetic variation in fall cold hardiness in coastal Douglas-fir in western Oregon and Washington. *Can J Bot* **84**: 1110–1121.
- THOMSON, A. J. and Y. A. EL-KASSABY (1988): Trend surface analysis of provenance-progeny transfer data. *Can J For Res* **18**: 515–520.
- VERBYLA, A. P., B. R. CULLIS, M. G. KENWARD and S. J. WELHAM (1999): The analysis of designed experiments and longitudinal data by using smoothing splines (with discussion). *Applied Statistics* **48**: 69–311.
- WAND, M. P. (2003): Smoothing and mixed models. *Comput Stat* **18**: 223–249.
- WOODS, J. H., D. KOLOTELO and A. D. YANCHUK (1995): Early selection of coastal Douglas-fir in a farm-field environment. *Silvae Genet* **44**: 178–186.
- WHITE, I. M. S., R. THOMPSON and S. BROTHERSTONE (1999): Genetic and environmental smoothing of lactation curves with cubic splines. *J Dairy Sci* **82**: 632–638.
- YE, T. Z. and K. J. S. JAYAWICKRAMA (2008): Efficiency of using spatial analysis in first-generation coastal Douglas-fir progeny tests in the US Pacific Northwest. *Tree Genet Genomics* **4**: 677–692.
- ZAS, R. (2006): Iterative kriging for removing spatial autocorrelation in analysis of forest genetic trials. *Tree Genet Genomics* **2**: 177–185.

## Clonal Variation in Lateral and Basal Rooting of *Populus* Irrigated with Landfill Leachate

By R. S. ZALESNY JR.<sup>\*)</sup> and J. A. ZALESNY

U.S. Forest Service, Northern Research Station, Institute for Applied Ecosystem Studies,  
5985 Highway K, Rhinelander, WI 54501, U.S.A.

(Received 21<sup>th</sup> January 2010)

### Abstract

Successful establishment and productivity of *Populus* depends upon adventitious rooting from: 1) *lateral roots* that develop from either preformed or induced primordia and 2) *basal roots* that differentiate from callus at the base of the cutting in response to wounding. Information is needed for phytotechnologies about the degree to which *Populus* adventitious rooting is controlled by effects of individual genotypes, waste waters used as alternative fertigation sources, and their interactions. Our objective was to irrigate twelve *Populus* clones with well water (control) or municipal solid waste landfill leachate and to test for differences between initiation of lateral versus basal roots, as well as root growth rate and distributional trends for both root types. We evalu-

ated number and length of lateral roots initiated from upper, middle, and lower thirds of the cutting, as well as basal callus roots. Overall, leachate irrigation affected lateral roots but not basal roots, and there was broad clonal variation between and within root types. On average, there were 129% more lateral than basal roots, which ranged from 3 to 27 (lateral) and 2 to 10 roots (basal). The percent advantage of number of roots from the middle portion of the cutting relative to other sections was 120% (upper), 193% (lower), and 24% (basal). Clones, treatments, and their interaction did not affect root growth rate, which ranged from  $1.5 \pm 0.6$  to  $3.4 \pm 0.3$  cm d<sup>-1</sup>, with a mean of  $2.3 \pm 0.2$  cm d<sup>-1</sup>. These results contribute baseline information for clonal selection needed to establish *Populus* for phytotechnologies, energy, and fiber.

**Key words:** forest genetics, tree improvement, phytotechnologies, hybrid poplar, root types, *Populus deltoides*, *P. trichocarpa*, *P. suaveolens* subsp. *maximowiczii*, *P. nigra*.

<sup>\*)</sup> Corresponding author (research plant geneticist):  
Phone: +1 (715) 362-1132, Fax: +1 (715) 362-1166.  
E-mail: [rzalesny@fs.fed.us](mailto:rzalesny@fs.fed.us)